

Geographic Distribution of Genetic Variation in the Hellbender, *Cryptobranchus alleganiensis*

Eric J. Routman¹

¹Department of Biology, San Francisco State University, 1600 Holloway Avenue, San Francisco, California 94132, USA

Summary

My students (Stephen Sabatino and Maria Tonione) and I investigated hellbender genetic variation through phylogenetic and population genetic analyses of individuals sampled from 17 locations throughout their range in the eastern United States. Mitochondrial DNA analyses were conducted on concatenated cytochrome-oxidase I (COI), cytochrome-b (Cytb) and NADH dehydrogenase subunit 4 (ND4) mtDNA sequences, totaling 2160 nucleotides. Most of the mtDNA results have been published (Sabatino and Routman 2008). Autosomal genetic variation was surveyed with a variety of single copy nuclear markers, although only four microsatellite loci were sufficiently variable for complete analysis.

Hellbender mtDNA haplotypes differed by 0.1% to 5.7% maximum likelihood corrected sequence divergence. Phylogenetic analyses reveal that hellbenders are separated into 8 reciprocally monophyletic groups differentiated by a minimum of 0.7% to 5.4% sequence divergence. Preliminary surveys on 8 autosomal markers revealed very low levels of DNA sequence variation, so further analysis was not possible. RAPD marker analysis revealed more variation, but individual RAPD profiles proved to be unrepeatable. Four microsatellite loci had sufficient variation for population genetic analyses. The average number of alleles per locus for the entire sample was 3.6. Unlike the situation with mtDNA, hellbender populations showed moderate among population sharing of alleles at microsatellite loci. However analysis using F-statistics and Bayesian clustering algorithms showed considerable population subdivision and clustered hellbender populations into the same major groups as the mtDNA.

High among-population divergence for microsatellites combined with reciprocal monophyly for mtDNA suggest that gene flow is severely restricted or non-existent among the 8 major groups, and potentially among populations (rivers) within group. Hellbenders are currently divided into two subspecies, *Cryptobranchus alleganiensis alleganiensis* and *C. a. bishopi* based on morphological characters. The analyses presented here indicate that each subspecies is probably paraphyletic and undoubtedly highly subdivided into separate units that are not genetically exchangeable. Management priorities for the hellbender should be reconsidered in light of these data.

INTRODUCTION

Phylogeography and Conservation

One persistent problem in evolution and conservation biology is the delineation of species boundaries. The terms "species" and "subspecies" are controversial, which has resulted in numerous definitions for each in the literature (Futuyma 1997). A common thread among several widely accepted definitions is the idea that species or subspecies are lineages that are evolving independently from other lineages. Genetic differentiation among groups of related organisms may be the best evidence of lineage separation as it is indicative of reproductive and/or ecological isolation and is the result of long-standing trends in demographic history (Avice 1995, 2000). The use of genetics is especially useful for identification of cryptic species or subspecies where morphological differentiation is absent or misleading (Baric and Sturmbauer 1999).

When geographically separate populations have been isolated from one another for many hundreds of thousands or millions of years, evolution within each population assures that sufficient morphological and genetic evidence will be present to identify each independent lineage using standard phylogenetic techniques. When the time frame of isolation is shorter, usually hundreds to tens of thousands of years, or when rare gene flow is present, phylogenetic techniques often fail to identify isolated lineages. In these situations, a combination of phylogenetic and population genetic approaches, using multiple genetic markers, can often reveal underlying structure in what appears superficially to be a single lineage.

Relatively independent populations, even those isolated for short time periods, are likely to possess unique local adaptations, because of the speed which selection can change populations. The preservation of locally adapted populations and maintenance of their genetic integrity is important since they are likely to be significantly contributing to the overall adaptive potential of the species (Lacy 1987; Templeton 1987). The genetic integrity of locally adapted populations can be compromised via management practices implemented to increase population sizes and/or genetic diversity of endangered species such as breeding programs or the translocation of individuals among currently isolated populations (Tallmon et al. 2004). This risk stems from the fact that adaptive phenotypes of individuals within independent populations are often controlled by unlinked multi-gene complexes (Wright 1968; Matisoff and Templeton 1999). If individuals from such genetically distinct populations are crossed, either through breeding programs or translocations, their hybrid offspring may receive portions of co-adapted gene complexes that are incompatible or unsuited to each other since they evolved in unique genetic backgrounds (i.e., outbreeding depression; Hedrick, 2001). Outbreeding depression leads to lower individual fitness (Hedrick 2001; Lenormand 2002; Gharrett et al. 1999) and can therefore hamper conservation efforts. Molecular phylogeography can therefore be used to identify which populations should be preserved and assess risks associated with conservation strategies that involve cross-breeding of independent populations.

The Hellbender

Hellbenders are large (up to 74 cm) aquatic salamanders found in clear fast-flowing streams in the eastern U.S. (Fig. 1). There are currently two recognized hellbender subspecies (Nickerson and Mays 1973). The Ozark hellbender (*Cryptobranchus alleganiensis bishopi*) inhabits streams that drain south out of the Ozark Plateau in the highlands of Missouri and Arkansas. All other populations of hellbenders, including those in rivers draining northward from

the Ozarks, belong to the nominal subspecies, *C. a. alleganiensis*. The Ozark hellbender is mainly characterized by its smaller spiracle size relative to the eastern hellbender, its coloration pattern and a few other minor morphological characteristics (Nickerson and Mays 1973). Hellbender populations are declining throughout most of their range (Nickerson et al. 2002; Wheeler et al. 2003), and are considered endangered or threatened by state wildlife agencies in most states where they are found (<http://www.hellbenders.org/conservation.html>). The reasons for the decline are uncertain and probably include habitat loss and disease (Wheeler et al. 2003). Captive breeding and release is now being considered (e.g. Ron Goellner Center for Hellbender Conservation at the St. Louis Zoo), and genetic population structure information is crucial to this enterprise.

The most recent investigation into hellbender population genetics, using restriction enzyme digestion of the entire mtDNA genome, found high levels of genetic differentiation (between .2 and 4% sequence difference) among most populations within each putative subspecies (Routman et al. 1994). The two named hellbender subspecies *C. a. alleganiensis* and *C. a. bishopi* were also shown to be paraphyletic. However, the relationships among major clades of mtDNA haplotypes were not well resolved

Here we explore the molecular phylogeography of the endangered hellbender salamander, *Cryptobranchus alleganiensis*, using mtDNA sequence data and a variety of autosomal gene markers. Our goals are to better understand the recent and historical relationships among hellbender populations, to test subspecies boundaries within the species, and to identify which hellbender populations may be particularly important to the long-term viability of the species. These data can be used to shape conservation management of hellbender populations.

Methods

Hellbender Sampling

Blood and/or tissue samples were collected from hellbenders caught in 17 different rivers and streams throughout the Appalachian and Ozark Mountains of the eastern US (Fig. 1 and Table 1; see Routman 1993 and Routman et al. 1994 for details). Tissue samples from Eleven Point River hellbenders were collected by Dr. Jeff Briggler of the Missouri Department of Conservation and stored in 95% ETOH. We extracted DNA from hellbender blood and tissue samples using the standard phenol/chloroform method outlined in Routman (1993). We further cleaned the DNA obtained from blood by adding 500ng of it to 100 µl of 5% Chelex (BioRad), boiling the mixture for 2 minutes and isolating the supernatant. All DNA samples were stored at -20°C.

Mitochondrial DNA

Polymerase Chain Reaction and Sequencing

Initially we used the Polymerase Chain Reaction (PCR) to amplify hellbender cytochrome-*b* (cytb) with a forward primer tRNA-Glutamine designed for *Dicamptodon* (see Table 2 for all primer sequences) and a reverse primer we designed from the consensus sequences of *Ambystoma laterale* (Genbank accession number: AY728227), *Desmognathus fuscus* (AY728218) and *Andrias davidianus* (NC_004926). The DNA template used for this particular PCR was extracted from hellbender ova, which contain a high amount of mtDNA relative to somatic tissue. The result was a 900bp fragment of cytb, which we sequenced and used to design the hellbender specific primers (Cytb-CA-F1, Cytb-CA-R1) that amplify a 750bp

fragment. We amplified cytochrome oxidase I (COI) using universal primers LCO 1490 and HCO 2198 (Folmer et al. 1994), which yielded a 500bp fragment. Because a larger fragment of COI is desirable, we aligned the resultant sequence with that of *Andrias davidianus* (NC_004926) and used it to design the primers COI-CA-F1 (based on *Andrias* sequence alone) and COI-CA-R1 (based on *Andrias/Cryptobranchus* alignment), which amplify a 750bp fragment. We amplified NADH dehydrogenase subunit 4 (ND4) with ND4-Nex-F and ND4-Nex-R from Hardy et al. (2002). These primers did not work for a number of hellbender samples, so we designed the hellbender specific primers ND4-Sal-F1 and ND4-Sal-R1 from *Andrias davidianus* (NC_004926) and hellbender sequence, which yielded a similarly sized fragment. Ampli-taq chemistry (Applied Biosystems (ABI)) was used for PCR under the following conditions: 94°C (120s) initial denature, {94°C (30s), 58°C (30s), 72°C (60s)} 30x, and a final extension at 72°C (120s). We sequenced DNA with ABI Big Dye chemistry on an ABI 377 sequencer according to manufacturer's recommendations.

We assessed within-population genetic diversity by conducting cytb, COI and ND4 allele counts. For every novel cytb or COI haplotype recovered, at least one individual that possessed it was also sequenced at ND4; in this way, we hoped to recover much of the mtDNA diversity in the hellbender populations studied and ensure there was sufficient statistical power for phylogenetic analyses.

Phylogenetic Analyses

We conducted maximum likelihood (ML) phylogenetic analyses on the concatenated cytb, COI and ND4 sequences. Redundant haplotypes were removed from the dataset prior to all phylogenetic analyses. We ran a maximum likelihood analysis (MLA), with and without *Andrias davidianus* sequence included as an outgroup, using the program **PAUP 4.0b10** (Swofford 1999). To choose a substitution model for MLA and Bayesian analysis (see below), we used the software program **Modeltest 3.7** (Posada and Crandall 1998). We conducted the following iterative MLA rather than a full one. We began by calculating the parameter values for the substitution model based on fixed tree topology generated via the bio-neighbor-joining algorithm in PAUP (Distance model = ML parameter estimates generated by ModelTest). The parameter estimates were then fixed and a heuristic ML search was run to find the most likely tree topology. The resultant tree was in turn used to re-estimate model parameter values. We continued the iterative process of estimating model parameters based on a fixed tree topology, and vice versa, until all model parameters, including the tree topology, stabilized. Independent estimation of model parameters and tree topology has been shown to accurately converge on the ML tree found when conducting a full MLA while dramatically increasing computational efficiency (Sullivan et al. 2005). We conducted bootstrap analysis using 500 ML bootstrap pseudoreplicates of the dataset, fixing model parameters at maximum likelihood values and estimating tree topology.

To root the hellbender mtDNA tree by enforcing a molecular clock, we first pruned the mtDNA dataset to include one representative haplotype from each of the 8 major monophyletic clades. The molecular clock hypothesis was tested for the data using a likelihood ratio test (Felsenstein 2004). We determined which evolutionary model to use to test the molecular clock hypothesis using Modeltest and conducted the phylogenetic analysis in PAUP.

The molecular clock rooting analysis was conducted using the program **MrBayes 3.1.2** (Huelsenbeck and Ronquist 2001). Since it is possible to specify separate evolutionary models for different parts of a dataset in MrBayes 3.1.2 and the 1st 2nd and 3rd nucleotide positions within

the protein-coding genes that make up the composite haplotypes are likely to be evolving at different rates, we used a separate model for each position. We determined which model to use for each position using the software program Modeltest. Bayesian trees were generated using MrBayes as follows: we ran two replicate searches, each of which was composed of four Metropolis-coupled chains and started with a random tree. Each chain ran for 1.0×10^7 generations and was sampled every 1000 generations. The chains were run until the average standard deviation of split frequencies among them reached .001, which yielded 10,000 trees from each run. The last 5000 trees generated in each run (for a combined total of 10,000) were used to determine the root. The Bayesian posterior probability (BPP) that a node is the root of the tree was calculated as the percentage of the 10,000 trees with that particular root. We calculated the BPP for each root position by importing all 10,000 Bayesian trees into PAUP and filtering them with constraint trees for each possible root placement on the 8 haplotype tree (see below).

Autosomal Genes

Autosomal Gene Sequencing

Because of the power of phylogenetic analysis to detect isolated lineages, our first choice for autosomal gene marker analysis was to sequence single copy genes, using the same basic methods employed for mtDNA. A total of eight nuclear introns were surveyed for variation in a preliminary subsample of individuals. Primers for one locus were obtained from the literature, and we developed primers for the other seven loci (Table 3). The primers were primarily developed from existing GenBank sequences, with the exception of the GATA locus. The primers for this locus were designed from a clone obtained from the microsatellite library used for the microsatellite analysis (see below). None of the 8 genes that were amplifiable possessed sufficient sequence variation for phylogenetic analysis (Table 4), so autosomal gene sequencing was not used to survey the entire sample.

RAPD Analysis

Randomly Amplified Polymorphic DNA (RAPD) can be used to detect variation among populations. This method uses PCR reactions primed with a single 10 bp oligonucleotide. Because the primer is so short, it randomly finds complementary matches at many different locations throughout the genome of the sample individual. Occasionally, these matching sites occur on opposite DNA strands in close enough proximity that the PCR reaction can amplify the segment. This results in a randomly amplified set of fragments for each individual and which may differ among individuals if there are differences in the sequence at the primer binding sites. We used the Ready-to-go RAPD analysis Kit from G. E. Healthcare to test several RAPD primers. However, although variation was detected, repeated reactions for several individuals yielded different fragment patterns for the same sample. Thus, the method was judged to be too unreliable for further analysis.

Microsatellites

Microsatellite loci are short sequences of simple repeats (e.g. CACACA) flanked by unique sequence DNA for which PCR primers can be developed, allowing the amplification of individual loci. These loci have been shown to be extremely common in eukaryotic genomes and variable in length, and thus are useful for population level studies. Thanks to our collaborators Dr. Jarrett Johnson (University of California, Davis) and Dr. Lori Eggert

(University of Missouri) we obtained primer sets for 4 microsatellite loci (Eggert et al in prep). These loci proved to be sufficiently variable for further analysis.

Three dinucleotide microsatellite loci (CRAL1, CRAL4 and CRAL9) and 1 tetranucleotide (GATA) were amplified using fluorescent PCR as described by Eggert et al (in prep). The forward primer was 5'-labelled with a fluorescent tag and the fluorescent product was then genotyped on an ABI 3100 sequencing machine (Applied Biosystems) and visualized using GeneScan Analysis Software (version 3.1; Applied Biosystems) with Filter Set D.

Data analysis

To estimate the frequency of null alleles and evidence of scoring errors due to stuttering in our dataset we used **Microchecker** version 2.2.3 (Van Oosterhout et al. 2004). Several descriptive genetic statistics were calculated on the populations, including the number of alleles, number of private alleles (alleles restricted to a single population), and expected heterozygosity (the probability of two alleles chosen at random from a population being different, a measure of genetic diversity) using the program **Arlequin** version 3.11 (Excoffier et al. 2005). We also used Arlequin to estimate observed (H_O) and expected (H_E) heterozygosities at each locus for each sample to test for deviations from Hardy-Weinberg equilibrium (HWE). Significance was evaluated using the Markov chain method (Markov chain parameters: 100,000 dememorization; 1,000,000 steps per chain) and we applied a Bonferroni correction (adjusted P value < 0.0009 , $\alpha = 0.05$). To test for linkage disequilibrium, we used **GENEPOP** version 4 (Raymond and Rousset 1995) with a Bonferroni correction (adjusted P value < 0.0008 , $\alpha = 0.05$).

We used **STRUCTURE** version 2.2 (Pritchard et al. 2000) to assign individuals to genetic clusters. This software uses a Bayesian algorithm that creates clusters assuming both Hardy-Weinberg equilibrium and gametic phase equilibrium within clusters and disequilibrium between clusters. The number of clusters (K) is set *a priori*, and the program fits the individuals to the K clusters with a Markov Chain Monte Carlo (MCMC) approach. Ten iterations were run for each value of K from $K=1$ (rivers constitute a single panmictic population) through $K=20$ (three more than the number of rivers sampled) using both the independent and admixture ancestry models, with a burn-in of 50 000 and MCMC chain lengths of 500000. We used the method described by Evanno et al. (2005) to identify the number of higher level groups of clusters in our dataset. Because locality data are not used in the analysis, the clusters provide an independent test of any hypotheses of population structure. The programs **CLUMPP 1.1.1** (Jakobsson and Rosenberg 2007). and **Distrupt 1.1** (Rosenberg 2004) were used to test the 10 replicates (for a given K value) within an analysis for compatibility, to combine replicates for a final estimate of individual membership probabilities, and to display the combined data graphically.

We used Arlequin to analyze population structure using Wright's **F-statistics** (Wright 1968), which are measures of the between group genetic variance, which is mathematically equivalent to the within group genetic covariance. The statistic F_{SC} measures the proportion of the total genetic variation attributable to the genetic differences among populations within clusters, averaged over clusters. The statistic F_{CT} measures the proportion of the total genetic variation attributable to differences among clusters. This analysis requires the populations and groups of populations be known *a priori*. Therefore we used the algorithm to test hypotheses of population relationships by clustering population samples into groups based on our mtDNA phylogeny, STRUCTURE clusters, or named subspecies and comparing the values of the F statistics.

Results

Mitochondrial DNA

Sequencing

A total of 26 cytb (N=88), 19 COI (N=91) and 17 ND4 (N=61) haplotypes were recovered from 16 hellbender populations (see Table 5). To ensure that mtDNA genes and not nuclear pseudogenes were sequenced, each sequence was translated, checked for stop codons and compared to the amino acid sequence of *Andrias davidianus* to ensure the reading frame was intact. Most individuals were sequenced at both cytb and COI, but this was not always the case due to variation in PCR and sequencing success. Concatenation of the three mtDNA gene sequences for individuals with complete data resulted in 32 unique mtDNA haplotypes that were 2160bp in length (807bp-cytb, 705bp-COI and 648bp-ND4) after sequence editing. Haplotypes from 6 individuals from the West Fork of the Greenbrier (a tributary of the New River) are not included in the phylogenetic analysis, because technical difficulties have delayed the sequencing of all three mitochondrial genes for these individuals. However, all 6 individuals have COI gene sequences that are identical to those from the New River, and 4 individuals have Cytb gene sequences that are identical or nearly identical to the New River sequences, which unambiguously places these samples in the New River clade (see below). We are continuing to sequence these samples in order to complete the data set.

The pairwise sequence divergences among these haplotypes, corrected for multiple substitutions using the maximum likelihood described below for the phylogenetic analysis, are shown in Table 6. They range from 0.1% to 5.7% sequence divergence. Considerable among-population genetic independence was observed, even at individual loci. Ten of 16 hellbender populations surveyed in this study share no 3-locus haplotypes with other populations. The 6 hellbender populations with shared haplotypes comprise the following 3 pairs that share haplotypes within pairs but not among them: (Big Piney River, Gasconade River), (Blue River, Slippery Rock Creek) (French Creek, Sherman Creek). The Big Piney River is a tributary of the Gasconade River, and these hellbenders could be considered a single population. The other four rivers contain haplotypes that are all closely related (see below). In contrast, haplotypes within each population appear closely related, differing by a maximum of 7 bases out of over 2000 bases sequenced.

Phylogenetic Analyses

Modeltest selected the TrN+G model to be the most appropriate substitution model for the mtDNA haplotype dataset with no outgroup using the Akaike Information Criterion (AIC). Maximum likelihood analysis resulted in the following model parameter estimates: $\Pi_A = 0.356$, $\Pi_C = 0.162$, $\Pi_G = 0.180$, $\Pi_T = 0.301$; AC = 1.000, AG = 20.979, AT = 1.000, CG = 1.000, CT = 26.965, GT = 1.000 and Γ -distribution shape parameter $\alpha = 0.161$. The midpoint rooted ML phylogeny (Fig. 2) has a likelihood score of $-\ln L = 4591.762$. The outgroup-rooted ML tree including the *Andrias* haplotype (tree not shown) has the same overall topology as the mid-point rooted tree except it places the Current River haplotype as a sister taxon to all others. However, because 20% sequence divergence exists between the *Andrias* haplotype and those of Current River hellbenders, *Andrias* is clearly a poor outgroup and the exact root placement is questionable. As a result, we used the Bayesian analysis to determine the root of the hellbender phylogeny.

The dataset used for Bayesian rooting included one haplotype from the following hellbender populations: Big Piney/Gasconade River, New River, Spring River, Current River, Little River, Slippery Rock Creek, Copper Creek and the North Fork of the White River. The results of Modeltest showed the GTR+G model to be most appropriate for the entire pruned dataset used to test the molecular clock hypothesis. The likelihood scores for the tree estimated with and without enforcing a molecular clock were 4238.54 and 4236.61, respectively. The likelihood scores for the two trees are not significantly different ($\chi^2 = 3.86$, $p = 0.70$, $df = 6$) and the molecular clock hypothesis cannot be rejected. Molecular clock rooting has been shown to be most accurate when sequences are evolving in a clock-like fashion and is therefore appropriate in this case.

The application of Modeltest indicated the GTR+G, Trn and the GTR+G models to be most appropriate for 1st, 2nd and 3rd codon positions, respectively, for the dataset used in the Bayesian molecular clock rooting analysis. Since it is not possible to specify the Trn model in MrBayes and Bayesian analysis is robust to slight over-parameterizations of models (Huelsenbeck et al. 2002), we used GTR for 2nd position nucleotides. Ten thousand trees were used to calculate the posterior probability of each possible root position in the 8-haplotype tree. About 92% of the sampled Bayesian trees generated while enforcing a molecular clock rooted the mtDNA tree on the long branch between the New River/Current River hellbender haplotypes and the rest of the tree (Figure 3). The root was found to be on the branch leading to the Current River hellbender haplotype or the New River haplotype in 3.8% and 4.1% of the trees, respectively. All other nodes were found to be the root less than 1.0% of the time. This result is completely consistent with the mid-point rooted tree.

The rooted ML phylogeny (Fig. 2) reveals hellbender mtDNA haplotypes are separated into 8 reciprocally monophyletic clades differentiated by 0.7% to 5.4% sequence divergence. The 8 groups, named by drainage (see Fig. 1, Table 1) or geographic region, are: 1) **Northern Ozarks**, 2) **Eleven Point/Current River**, 3) **North Fork of the White**, 4) **Spring River**, 5) **New River**, 6) **Copper Creek**, 7) **Tennessee River**, and 8) **Ohio/Susquehanna River**. Moderate (70%-80%) to strong (above 90%) ML bootstrap support was found for the placement of all clades separated by more than 0.7% sequence divergence. In general, resolution at the tips of clades is poor, except for the Meramec River clades, whose monophyly is supported by a bootstrap frequency of 87%, and a subset of the Ohio-Susquehanna Rivers clade (82% bootstrap support).

Microsatellite Analysis

Allelic variation

The number of alleles per microsatellite locus in our samples ranged from 1 to 10, with a mean of 3.6. (Table 7). Expected heterozygosities (equivalent to the Hardy-Weinberg expected value adjusted for sampling) for each population are shown in Table 8. The average expected heterozygosity over all loci and populations was 0.555, a high value for protein-coding loci but normal for the hypervariable microsatellites. Genotype frequencies conformed to Hardy-Weinberg in all populations across all loci, after Bonferroni correction of probability levels for multiple comparisons. There was no evidence of allele stutter or null alleles.

Population structure

The results of the STRUCTURE analysis, using a model of independent populations, estimated the most probable value of K to be 11 clusters (Bayesian posterior probability =

0.954). However, when populations are distributed in hierarchical groups, in which migration among populations within a group is greater than migration among groups, Evanno et al. (2005) showed that the most likely value of K does not provide an accurate estimate of the number of higher level groups. To estimate the number of population groups, we employed the method outlined in Evanno et al. (2005) which uses the variance-adjusted rate of change in the likelihood of K (ΔK) to estimate the number of clusters. Using this approach, the highest ΔK was found at $K = 12$, with another peak at $K=8$ (Figure 4.)

Examination of the graphical output from STRUCTURE shows a high level of concordance between the clustering of the hellbender microsatellite loci and the geographic clustering of mtDNA haplotypes. (Figs. 5, 6, and 7). Figure 5 compares the STRUCTURE results for $K=8$ with the phylogeny of the mtDNA. The colors in the bar chart represent the different genetic clusters. Each bar represents an individual, and the proportional representation of the colors in each bar represents the probability of that individual being a member of that cluster. The 8 mtDNA clades are indicated at the top of the bar chart. Note that the most likely color (cluster) for each of the 8 groups is unique, with a few individual exceptions (several of these are due to missing data at one or more loci, which can make cluster assignment difficult) and that the New River Clade hellbenders appear to be similar to the Copper Creek hellbenders. Given that the clusters were created without regard to the geographic location from which the individual was sampled, the concordance of the clusters with geography and with the 8 mtDNA clades suggests real genetic isolation among these 8 regions. The presence of multiple clusters within each region may be due to substructuring within regions that is not fully detectable with our sample sizes (4 loci, 2-24 individuals per sample). Average cluster membership for each population is given in Appendix 2.

Figure 6 compares the STRUCTURE output for $K=8$ and $K=11$ (the most probable number of clusters). [Colors cannot be compared between different graphs, because the graphs represent independent analyses and the program assigns color randomly.] Although with additional clusters individuals often have greater ambiguity as to cluster membership, the basic geographic division into 8 groups is the same, again with the exception of a few individuals. Figure 7 compares $K=11$ to $K=12$ (the highest peak in the Evanno analysis). Again the basic geographic pattern holds, except that the two analyses seem to differ about the relationship of the New River hellbenders to the other rivers. In the $K=11$ analysis, the New River group is most similar to the Copper Creek group, but in the $K=12$ analysis, the New River group is most similar to the Ohio/Susquehanna group or to the hellbenders from Beaverdam Creek. There is no way to test this contrast statistically however.

We used Wright's F -statistics to explore how the hellbender microsatellite variation was distributed among possible clustering of the samples. The statistic F_{SC} measures the proportion of the total genetic variation attributable to the genetic differences among populations within clusters, averaged over clusters. The statistic F_{CT} measures the proportion of the total genetic variation attributable to differences among clusters. When the microsatellite samples were clustered into 8 groups matching the mtDNA clades, $F_{CT} = .377$ and $F_{SC} = .080$, indicating that populations within clusters are similar but populations between clusters are genetically different. When the samples were clustered into two groups matching the subspecies, $F_{CT} = .151$ and $F_{SC} = .354$, suggesting that the subspecies comprise populations that are more different from one another than are populations from different subspecies.

Discussion

If we base our conclusions on mtDNA alone, most hellbender populations surveyed in this study are genetically distinct. Mutually exclusive mtDNA haplotypes for most populations surveyed and mutually exclusive haplotype distribution among the 8 major monophyletic clades suggest that migration among hellbender populations, at least by females, is restricted or nonexistent.

Hellbender populations from within each of the northern Ozarks, Ohio/Susquehanna Rivers and Tennessee River regions are genetically similar. The sequence divergence among populations within each of these regions is less than 0.3%. This is an expected result given the populations are geographically associated and share river drainages. However, our data do suggest that some population genetic structure may exist even within these regions. In the northern Ozarks region, the Meramec River enters directly into the Mississippi River while the Big Piney, Gasconade and Niangua Rivers are tributaries of the Missouri River (Table 1, Fig. 1). Monophyly of Meramec River mtDNA haplotypes suggests the Mississippi River and/or the Missouri River are barriers to gene flow. Hellbender haplotypes from the Little River, which is a tributary of the Tennessee River, also form a monophyletic clade, although with weak branch support. In the Ohio/Susquehanna Rivers region, Sherman Creek hellbenders share a haplotype with those in French Creek while Slippery Rock Creek individuals share a haplotype with Blue River individuals, but haplotypes found in all four of these populations are closely related.

The Sherman Creek hellbender population is the only one sampled from a waterway that does not ultimately flow into the Mississippi River (it drains into the Susquehanna River, see Table 1, Fig.1). Therefore, at present Sherman Creek hellbenders are completely isolated from all other sampled populations. The only Sherman Creek hellbender haplotype found in this study is shared by individuals within the Ohio River populations (French Creek), indicating the migration of hellbenders into the Susquehanna or vice versa was relatively recent. Colonization may have occurred via stream capture or a recent connection between the two drainages (Routman et al. 1994). Since the Susquehanna River is presently isolated and the rooted mtDNA tree shows Sherman Creek hellbenders as one of the youngest lineages, it is reasonable to assume it was colonized by one of the Ohio River populations.

Surprisingly, there is relatively little genetic divergence (sequence divergence $\approx 0.7\%$) between the haplotypes from hellbenders found in rivers flowing north out of the northern Ozarks into the Missouri and Mississippi Rivers and those found in the Ohio and Susquehanna River tributaries (excluding the New River). This shallow divergence indicates a recent gene flow connection between these regions, compared to other hellbender populations (see also Routman 1994). The relatively close relationship between the northern Ozark and Ohio/Susquehanna River hellbender populations sharply contrasts with relationships among rivers in the southern Ozarks, which contain populations that are extremely divergent from one another despite sharing the White River drainage and being separated by tens of river miles, as opposed to hundreds. For example, hellbenders in the North Fork of the White River differ from those in the Spring River and Current River by 1.7% and 5.3% sequence divergence, respectively. Astonishingly, despite the fact that the Spring, Current and Eleven Point Rivers drain into the same tributary of the White River (the Black River) at almost the same point, hellbenders from the Spring River are much more closely related to those in the North Fork of the White River. These data suggest that long-standing barriers to gene flow exist in parts of the southern Ozarks while gene flow barriers dividing the much larger geographic area of the

northern Ozarks and the Ohio River are considerably younger or weaker. These remarkable contrasting results exemplify how the evolutionary history of hellbender populations varies by location, probably due to dissimilar geologic histories.

The relationship of Copper Creek hellbenders as a sister taxon to the [Northern Ozark, Ohio/Susquehanna River, Tennessee River] clade is moderately supported by maximum likelihood analysis. However, all our analyses (including parsimony and neighbor joining distance analyses, data not shown) agree on this relationship. Routman et al. (1994) found a sister taxon relationship between Copper Creek hellbenders and those in Tennessee River populations, which seems reasonable from a geographic perspective and is not in conflict with our findings. However, given that statistical support of this relationship is not strong in either study, additional research is needed.

The placement of the [Current River/Eleven Point River, New River] clade as sister taxon to the rest of the hellbender haplotypes is consistent with the hypothesis that these rivers, or other southern rivers to which they were connected, served as Pleistocene glacial refugia. Pleistocene glaciations created the Ohio River (Ray 1974). Therefore, during the Pleistocene, glaciers must have extirpated most aquatic organisms inhabiting rivers at the same latitude or higher than the Ohio River currently sits (Pielou 1991). Southern Ozark, Tennessee River and New River populations, on the other hand, are located south of the latitude to which glaciers reached during the last ice age and therefore pre-Pleistocene relationships among those populations may have persisted through the Pleistocene glaciations. Moreover, southern Ozark hellbender populations and those in the vicinity of the Tennessee River would have been unaffected by glacial runoff and are the most likely source of post-Pleistocene recolonization (Routman 1994). Thus the phylogeographic patterns observed in this study are consistent with reasonable expectations of how Pleistocene glaciations would have affected the evolutionary history of hellbenders.

The relationship between New River and C/EP Rivers hellbenders is unexpected and does not indicate ongoing gene flow. It is likely that this relationship reflects an ancient colonization event or the chance retention of the same part of an ancestral polymorphism independently by both populations. This hypothesis is supported by a comparable phylogeographic pattern found in the northern hogsucker (*Hypentelium nigricans*), in which individuals from the New River were most closely related to those in the northern and southern Ozarks (Berendzen et al. 2003).

Unfortunately, although we sequenced 3 mitochondrial genes, our mitochondrial data comprise a single genetic marker because mtDNA lacks recombination among genes. A single genetic marker is insufficient for drawing strong phylogeographic conclusions because of evolutionary idiosyncrasies that may have caused the geographic pattern of variation. Mitochondrial DNA, which is inherited cytoplasmically from only from the maternal parent and is effectively haploid, is though to be especially likely to display a geographic pattern of variation that is different from the rest of the genome. This is because male mediated gene flow does not transmit mtDNA, and because cytoplasmic inheritance allows mtDNA to more easily introgress into other species after hybridization (because mitochondrial genes are unlinked to genes in the nucleus responsible for reproductive isolation among species) (Wilson and Bernatchez 1998). For example, if among-river migrants in hellbenders were common but exclusively male, mtDNA would show complete isolation while the nuclear genome would be completely panmictic.

Including additional, autosomal gene markers helps to determine whether patterns of genetic variation reflect the evolutionary histories of the populations. Sequence data would be

the best to compare with the mtDNA results, because comparison of phylogenies is a relatively straightforward process. However, insufficient variation was detected to allow phylogeny estimation. The low levels of autosomal intron DNA sequence among hellbender populations is not unexpected, because mtDNA has a higher rate of evolutionary change compared to most nuclear DNA. This lack of variation does not mean that the mtDNA results are misleading.

The relatively few population-specific microsatellite alleles could be taken as an indication that gene flow among populations is high and that the populations are not evolutionarily isolated as suggested by the mtDNA phylogeny. I believe that the population genetic analysis of these loci suggests otherwise. The STRUCTURE and F-statistics results confirm the conclusion that the two subspecies are probably not monophyletic. The fact that STRUCTURE analysis defines clusters (or sets of clusters) that match well with the eight major mitochondrial clades, despite the fact that the clusters are created without geographic information, also strongly suggests that at least these eight groups of populations comprise evolutionarily distinct units. Even STRUCTURE analyses with $K > 8$, such as $K = 11$ or $K = 12$ are fairly consistent with the 8 mitochondrial clades, and the higher K values could be explained by additional subdivision within the 8 clades.

The sharing of microsatellite alleles by populations from different clusters could be explained by low but non-zero levels of gene flow, but the complete reciprocal monophyly and high levels of sequence divergence for mtDNA would require that gene flow be strongly male biased. Mark-recapture studies show low within-river movement and remarkable philopatry for both genders in adult hellbenders (Nickerson and Mays 1973; Peterson 1987; Routman unpublished data from the Meramec River), and larvae, which are more likely than adults to move long distances by being washed downstream, are unlikely to exhibit male bias.

More plausible explanations for microsatellite allele sharing in the presence of mitochondrial reciprocal monophyly may be found in the molecular biology and population genetics properties of microsatellites and mtDNA. One possibility is that, because microsatellites are prone to homoplasy caused by independent mutation to alleles of the same size (Culver et al. 2001), some of the shared alleles may have evolved independently in different populations. Even more likely, in my opinion, is that the sharing of alleles may represent the retention of ancestral polymorphisms for microsatellites but not for mtDNA. Coalescence theory (Kingman 1982) shows that the average time to the common ancestor for a sample of autosomal genes is approximately $4N_e$ generations, where N_e is the inbreeding effective population size of the population. This means that for populations to be reciprocally monophyletic they must be separated from each other for at least $4N_e$ generations, on average. When populations have been separated for shorter periods of time, they will possess alleles that are identical to (or most closely related to) alleles from a different population. However, the average time to the common ancestor for a sample of mitochondrial genes is approximately N_e generations, because maternal inheritance and haploidy results in 1/4 the number of copies of the mtDNA compared to autosomal genes (assuming a 1:1 sex ratio). This means that completely isolated populations will still be sharing autosomal alleles long after mitochondrial genes have become completely different. This scenario can be explored for our data using coalescent computer simulations, but because of the large number of populations and the wide range of possible population sizes and population divergence times, we did not attempt this in the course of our 1 year study.

Conclusions and Management Recommendations

Our data suggest that populations of *Cryptobranchus alleganiensis* comprise at least 8 distinct genetic units: 1) **Northern Ozarks**, 2) **Eleven Point/Current River**, 3) **North Fork of the White**, 4) **Spring River**, 5) **New River**, 6) **Copper Creek**, 7) **Tennessee River**, and 8) **Ohio/Susquehanna River**. This may be an underestimate of the population subdivision, because there is some evidence for substructure within some of the 8 clades (e.g. the Meramec River clade within the Northern Ozarks and the Little River clade within the Tennessee River group). Although some of the STRUCTURE analyses indicate that some of these 8 clades may be mixed, or that the New River hellbenders may be similar to the Tennessee River or Copper Creek hellbenders, I believe that the mtDNA results more accurately reflect population subdivision. The sequence-based phylogenetic analyses used for the mtDNA require smaller sample sizes for statistically significant results than the frequency-based microsatellite analyses. Further STRUCTURE analyses with additional loci and/or additional hellbender samples from some of the rivers would be expected to "clean up" some of our results, and perhaps detect some of the within clade substructure. (We have developed an additional microsatellite locus, and are in the process of surveying the samples, but we will not be finished in time for this report).

Our data also suggest that the two named subspecies are paraphyletic. Even if the phylogenetic analyses are in error about the relationships, and the subspecies are monophyletic, it is undeniably true that not all members of a subspecies are "created equal". For example, hellbenders from the North Fork of the White River are not genetically equivalent to hellbenders from the Spring River or the Current River, even though they are all *C. a. bishopi*.

The fact that even with relatively small sample sizes the STRUCTURE analyses confirm the existence of the 8 groups of hellbenders shows that these 8 groups have had time to evolve differences in these (presumably) selectively neutral markers. Since it takes even less time for selection to create differences among isolated populations, these groups are likely to have evolved local adaptations to their environments. This means that the 8 groups should be managed independently. This will prevent mixing of populations and the disruption of coadapted gene complexes and possible outbreeding depression.

Hellbenders from one group should not be moved into rivers containing hellbenders from another group. However, these 8 groups are almost certainly an underestimate of the real number of genetically distinct entities within the species, because the relatively small sample sizes also mean that some subdivision is being missed due to lack of power. Therefore, I recommend that each individual river be considered a management unit, unless circumstances on the ground prohibit such a fine scale approach. Transplants of individuals, or captive breeding programs, should preferentially be done within river, or, to reintroduce hellbenders to a river from which they have been extirpated, a source population from the nearest population from the same group should be used.

The exception to this recommendation would be in the case of a range-wide extinction catastrophe. If hellbender population decline becomes so precipitous that captive breeding is required to save the entire species or a large part of the species, I would recommend intentionally mixing individuals from different groups. This will provide the new population a large amount of genetic variation on which selection may operate. Although some individuals may suffer from outbreeding depression, hopefully the increased genetic variation will allow the evolution of new adaptive gene complexes. This approach has been tried in the reintroduction of collared

lizards to glades in the Missouri Ozarks (Templeton et al. 1990). Because the source populations (other Ozark glades) had low population sizes themselves, the only option was to choose one or a few lizards from each of many (genetically different) glades. These reintroductions proved to be very successful, indicating that outbreeding depression was overcome by selection on the new and highly variable founder population. A similar approach may prove useful for the hellbender.

Acknowledgements

Sequencing was done in the Conservation Genetics Laboratory at San Francisco State University. We thank Jeff Briggler from the Missouri Department of Conservation, who kindly provided specimens from the Eleven Point River. I also thank Victoria Grant, who made the funding process as painless as possible.

References

- Avise JC (1995) Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. *Conserv Biol* 9: 686-690.
- Avise JC (2000) *Phylogeography*. Harvard University Press, Massachusetts.
- Baric S, Sturmbauer C (1999) Ecological parallelism and cryptic species in the genus *Ophiothrix* derived from mitochondrial DNA sequences. *Mol Phylo Evo* 11:157–162
- Berendzen PB, Simons AM, Wood RM (2003) Phylogeography of the northern hogsucker, *Hypentelium nigricans* (Teleostei: Cypriniformes): genetic evidence for the existence of the ancient Teays River. *J Biogeogr* 30: 1139 -1152.
- Culver M, Menotti-Raymond MA, Obrien SJ (2001) Patterns of size homoplasy at 10 microsatellite loci in pumas (*Puma concolor*). *Mol Biol Evol* 18:1151-1156
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 14:2611-20
- Excoffier L, Laval G and Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47-50
- Felsenstein J (2004) *Inferring Phylogenies*. Sinauer, Massachusetts.
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek RC (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Bio Biotech* 3: 294-299
- Futuyma DJ (1997) *Evolutionary Biology*. Sinauer Associates, Massachusetts
- Gharrett, AJ, Smoker WW, Reisenbichler RR, Taylor SG (1999) Outbreeding depression in hybrids between odd- and even-brookyear pink salmon. *Aquaculture* 173: 117-129
- Hardy ME, Grady JM, Routman EJ (2002) Intraspecific phylogeography of the slender madtom: the complex evolutionary history of the Central Highlands of the United States. *Mol Eco* 11: 2393-2403
- Hedrick P (2001) Conservation genetics: where are we? *TREE* 16; 629-636
- Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17: 754-755
- Huelsenbeck JP, Bollback JP, Levine AM (2002) Inferring the root of a phylogenetic tree. *Syst Bio* 51:32-43
- Jakobsson, M. and Rosenberg, N. A. (2007). *CLUMPP*: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801-1806
- Kingman J.F.C. (1982) On the genealogy of large populations. *J Applied Probability* 19A:27-43
- Lacy RC (1987) Loss of genetic diversity from managed populations: Interacting effects of drift, mutation, selection, and population subdivision. *Cons Bio* 2: 143-158
- Lenormand T (2002) Gene flow and the limits to natural selection. *Trends Ecol Evol* 17: 183-189
- Matoli SR, Templeton AR (1999) Coadapted gene complexes for morphological traits in *Drosophila mercatorum*. Two-loci interactions *Heredity* 83: 54-61
- Nickerson MA, Mays CE (1973) A study of the Ozark hellbender, *Cryptobranchus alleganiensis bishopi*. *Ecology* 54:1163–1165
- Nickerson MA, Krysko KL, Owen RD (2002) Ecological status of the hellbender (*Cryptobranchus alleganiensis*) and the mudpuppy (*Necturus maculosus*) salamanders in the Great Smoky Mountains National Park. *J N Car Acad Sci* 118:27-34

- Peterson CL (1987) Movement and catchability of the hellbender, *Cryptobranchus alleganiensis*. J Herpetology 21: 197-204
- Pielou EC (1991) After the ice age – The return of life to glaciated North America. The University of Chicago Press, Illinois
- Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14: 817-818
- Ray LL (1974) Geomorphology and Quaternary geology of the glaciated Ohio River Valley-a reconnaissance study. U.S. Geological Survey Professional Publications. 826: 1-7
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J. Heredity 86:248-249
- Rosenberg, N. A. (2004). Distruct: a program for the graphical display of population structure. Molecular Ecology Notes 4: 137-138
- Routman EJ (1993) Mitochondrial DNA variation in *Cryptobranchus alleganiensis*, a salamander with extremely low allozyme diversity. Copeia 1993: 375-384
- Routman EJ, Wu R, Templeton AR (1994) Parsimony, molecular evolution, and biogeography: the case of the North American giant salamander. Evolution 48: 1799-1809
- Sabatino SJ and Routman EJ (2008) Phylogeography and conservation genetics of the hellbender salamander (*Cryptobranchus alleganiensis*). Conservation Genetics. *in press*
- Steele CA, Carstens BC, Storfer A, Sullivan J (2005) Testing hypothesis of speciation timing in *Dicamptodon copei* and *Dicamptodon aterrimus* (Caudata: Dicamptodontidae). Mol Phylo Evo 36: 90-100
- Sullivan J, Abdo Z, Joyce P, Swofford DL (2005) Comparing successive approximations and simultaneous optimization approaches to maximum likelihood estimation of phylogeny from DNA sequences. Mol Biol Evol 22:1386-1392
- Swofford DL (2001) PAUP* Phylogenetic analysis using parsimony (*and other methods). Version 4.0b8. Sinauer, Massachusetts
- Tallmon DA, Luikart G, Waples RS (2004) The alluring simplicity and complex reality of genetic rescue. Trends Ecol Evol 19: 489-496
- Templeton AR (1987) Coadaptation and outbreeding depression. In: ME Soule (ed) Conservation biology: The science of scarcity and diversity. Sinauer Associates, Massachusetts, p 105-116
- Templeton, A.R., K. Shaw, E. Routman, and S. Davis 1990. The genetic consequences of habitat fragmentation. Annals of the Missouri Botanical Garden. 77:13-27.
- VanOosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4: 535–538
- Wilson CC, Bernatchez L (1998) The ghost of hybrids past: fixation of arctic charr (*Salvelinus alpinus*) mitochondrial DNA in an introgressed population of lake trout (*S. namaycush*). Mol Eco 7: 127-132
- Wright S (1968) “Evolution and the Genetic of Populations” University of Chicago Press, Illinois
- Wheeler BA, Prosen E, Mathis A, Wilkinson RF (2003) Population declines of a long-lived salamander: a 20+ year study of hellbenders, *Cryptobranchus alleganiensis*. Biol Cons 109: 151-156

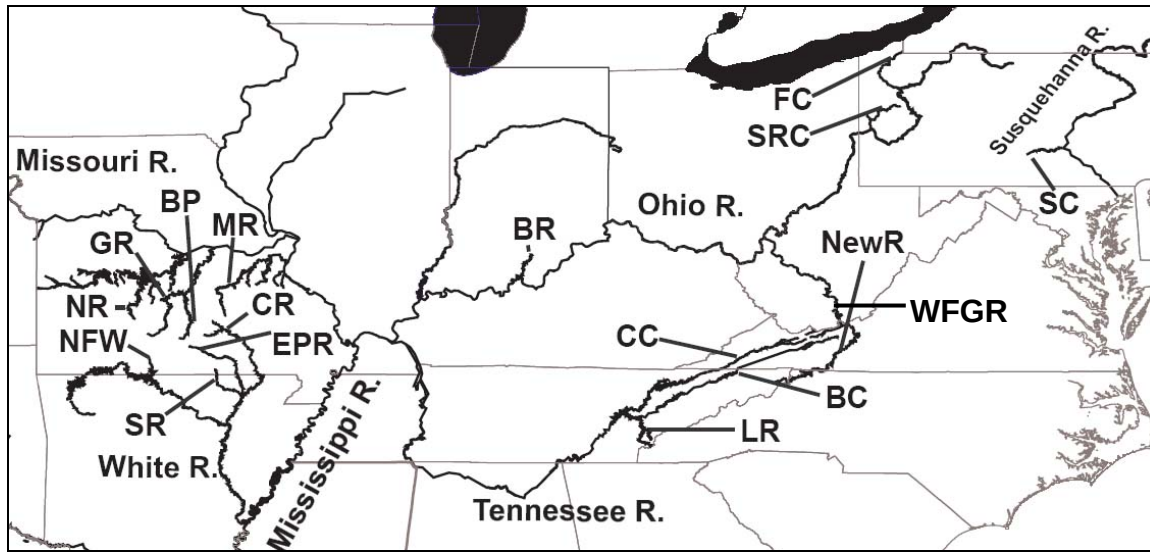


Figure 1. Map of the eastern USA showing the locations where hellbenders were sampled for this study. Major rivers are spelled out. Rivers where hellbenders were sampled are abbreviated as follows: Gasconade River (**GR**), Big Piney River (**BP**), Niangua River (**NR**), Meramec River (**MR**), Little River (**LR**), Beaverdam Creek (**BC**), Copper Creek (**CC**), Blue River (**BR**), French Creek (**FC**), Slippery Rock Creek (**SRC**), New River (**NR**), West Fork of the Greenbrier River(**WFG**), Sherman Creek (**SC**), North Fork of the White River (**NFW**), Spring River (**SR**), Current River(**CR**), Eleven Point River (**EPR**). The last four populations represent the subspecies *C. a. bishopi*. All other populations represent *C. a. alleganiensis*.

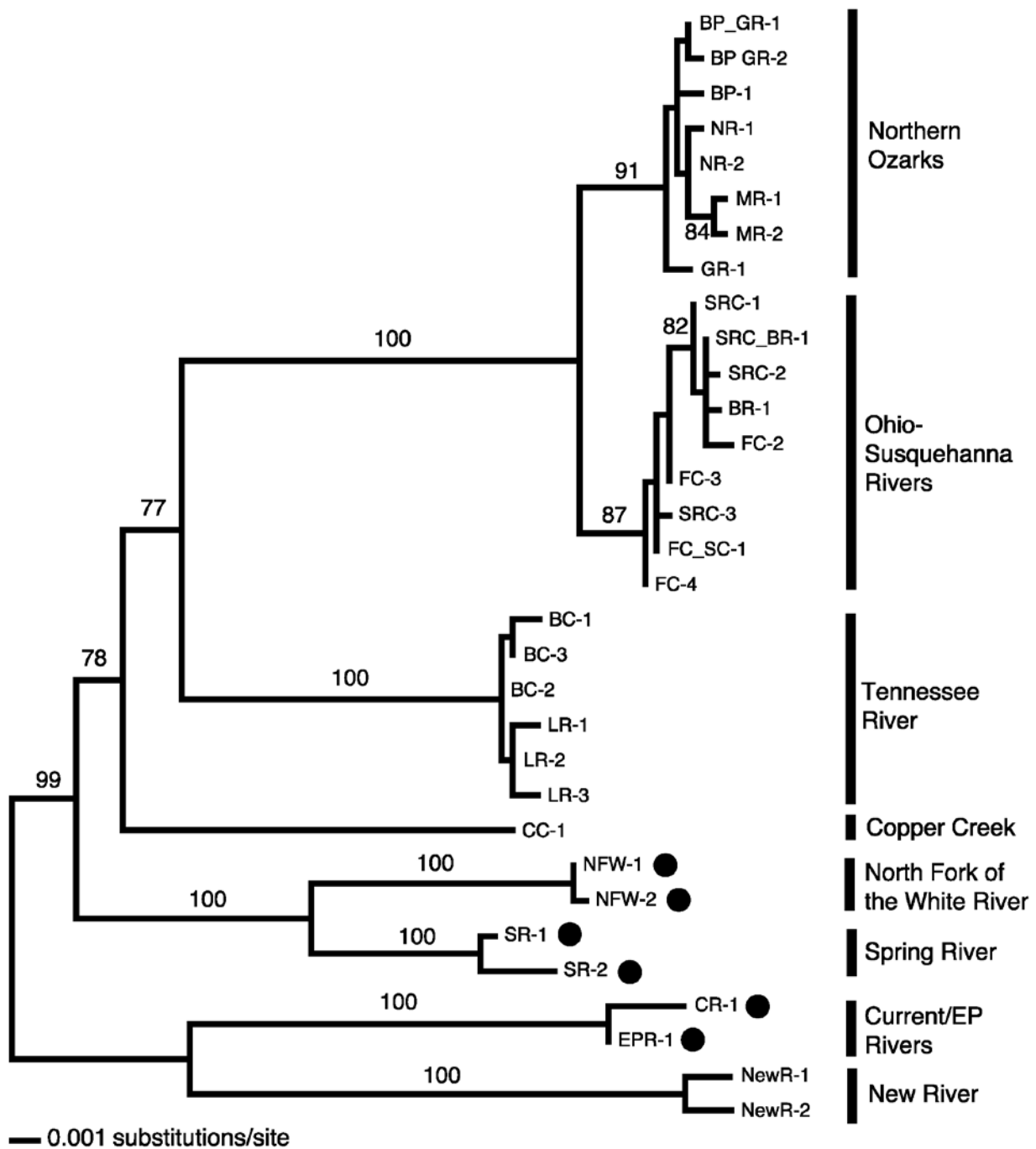


Figure 2. Midpoint-rooted maximum likelihood (ML) phylogeny of hellbender mtDNA 3-locus haplotypes. The numbers on branches represent ML bootstrap values (500 replicates). Haplotypes are labeled with letters indicating the river in which they were found and a number to separate different haplotypes in the same river. Haplotypes found in more than one river are given both river names. More inclusive clades are named by major drainage or geographic region. Haplotypes followed by a black circle were those found in hellbenders currently designated *C. a. bishopi*; all other haplotypes are considered to be *C. a. alleganiensis*.

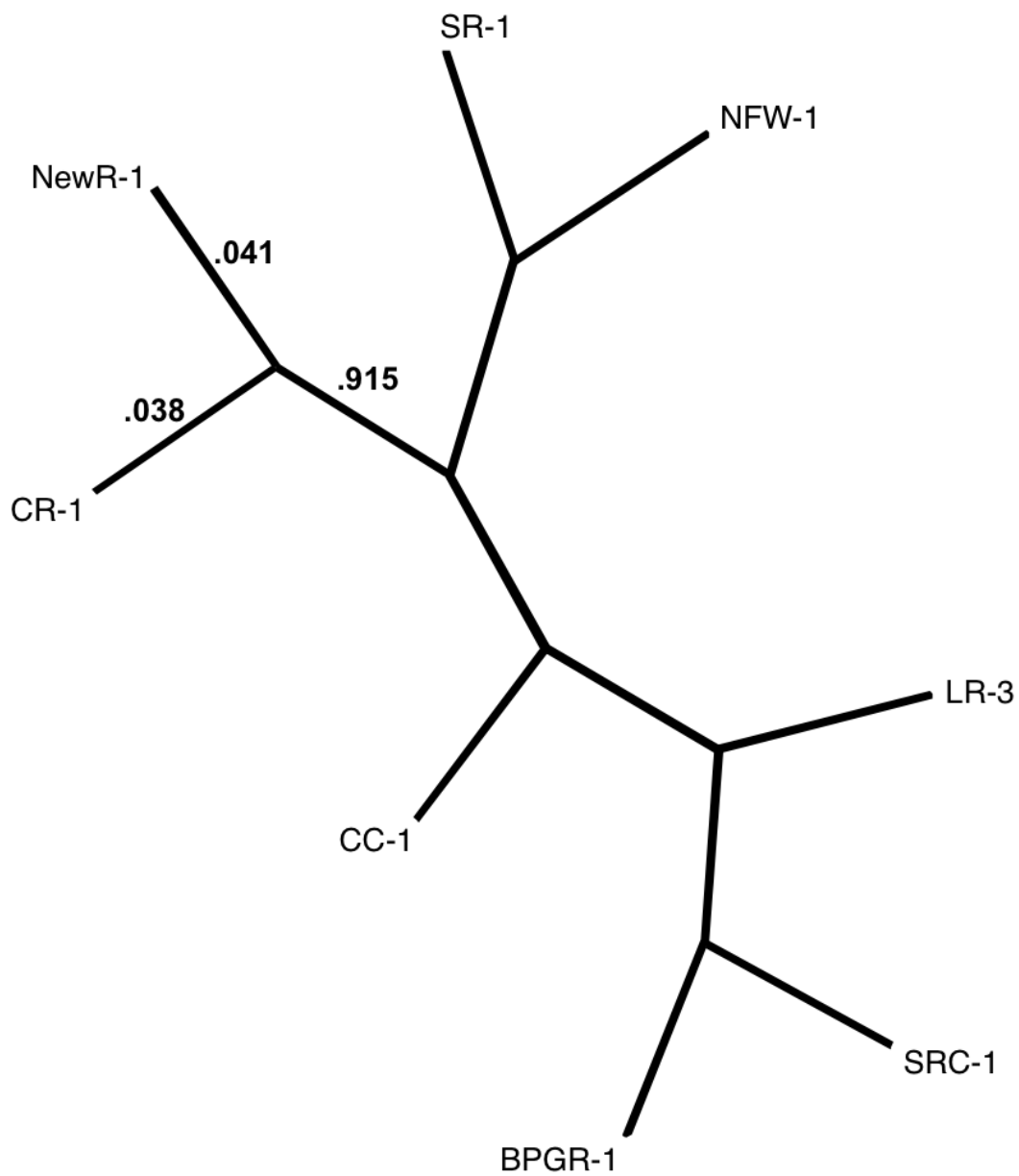


Figure 3. One of the constraint trees used to filter the Bayesian trees generated under a molecular clock. The branch leading to the New River and Current River clade was found to be the root in 91.5% of the trees sampled from the posterior distribution.

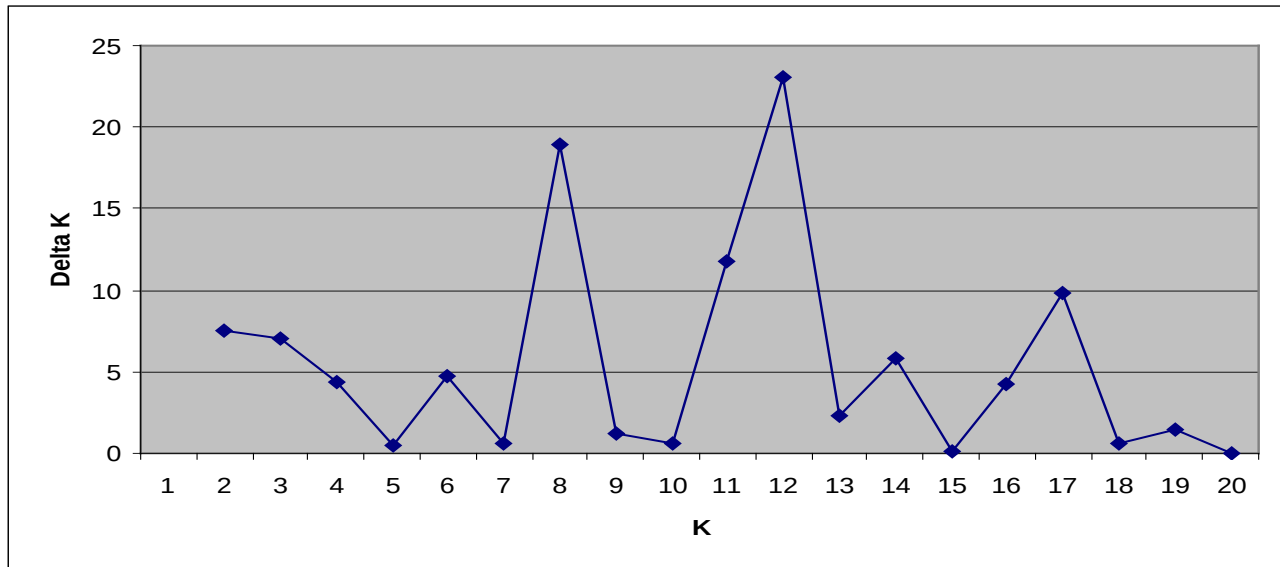


Figure 4. Plot of ΔK for the microsatellite data. Peaks at 8 and 12 suggest that populations are hierarchically grouped into 8 or 12 groups.

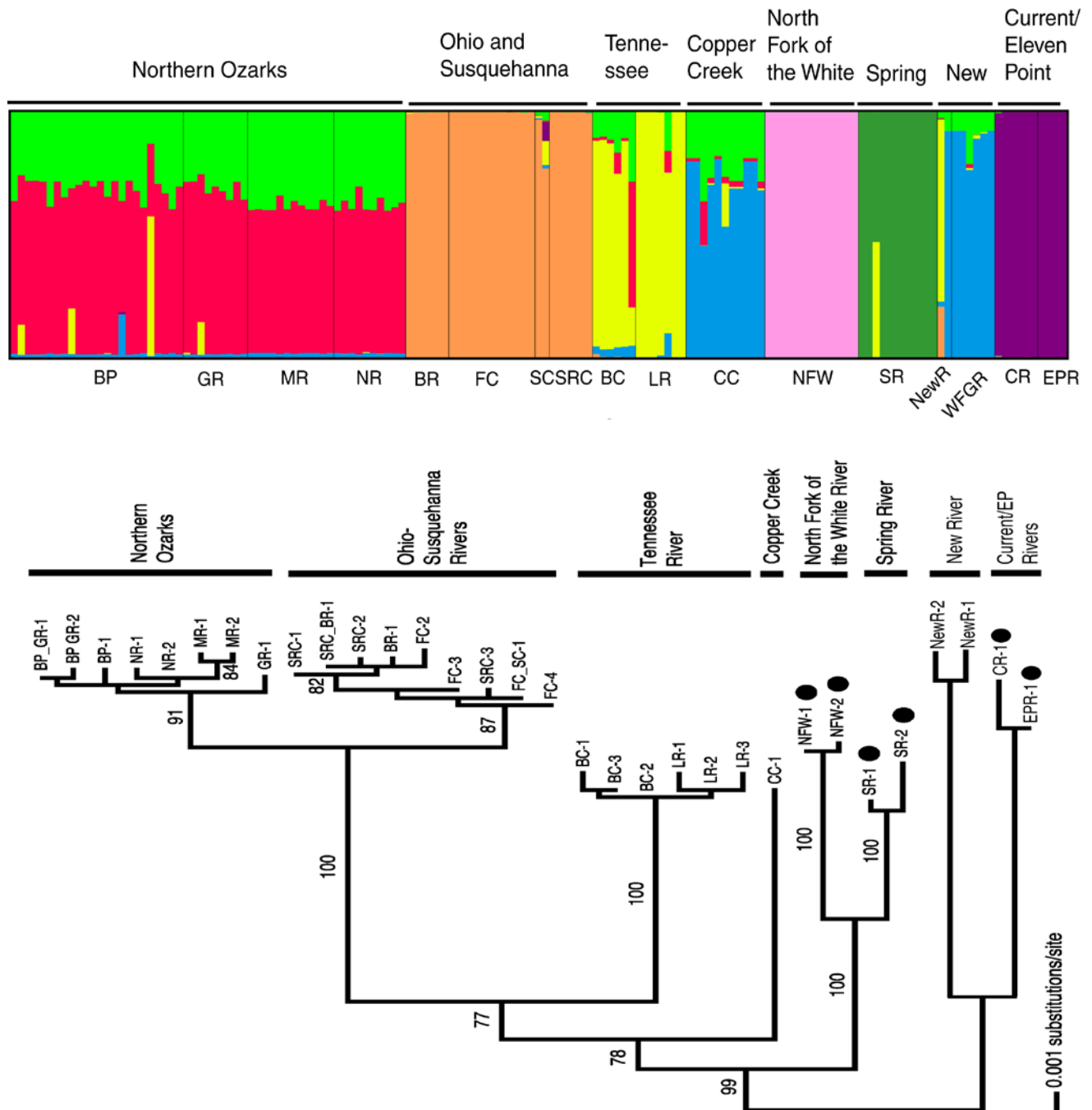


Figure 5. A comparison of the clusters estimated with STRUCTURE for K=8 with the phylogeny of mtDNA haplotypes (from Fig. 2.) Both analyses suggest that hellbenders populations are divided into 8 groups. Colors in the STRUCTURE output represent different genetic clusters. Bars on the chart represent each individual, with the proportion of each color equaling the probability of that individual's membership in that cluster. Individuals are arranged by population, with population abbreviation below and populations separated by black lines.

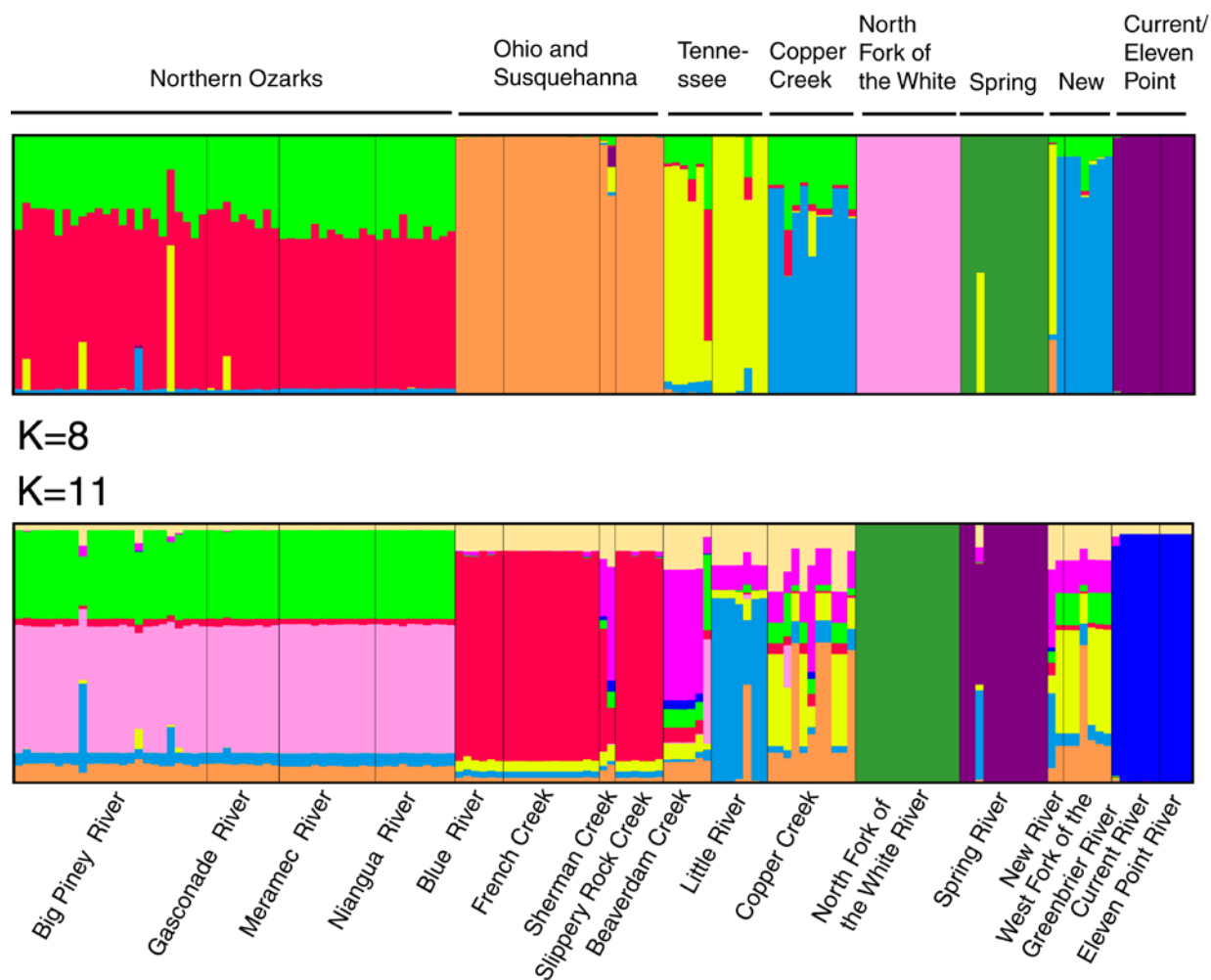


Figure 6. A comparison of the clusters estimated with STRUCTURE for K=8 vs. K = 11. Interpretation of the charts is as in Fig. 5. Colors cannot be compared across charts, because color assignment is independent in each analysis.

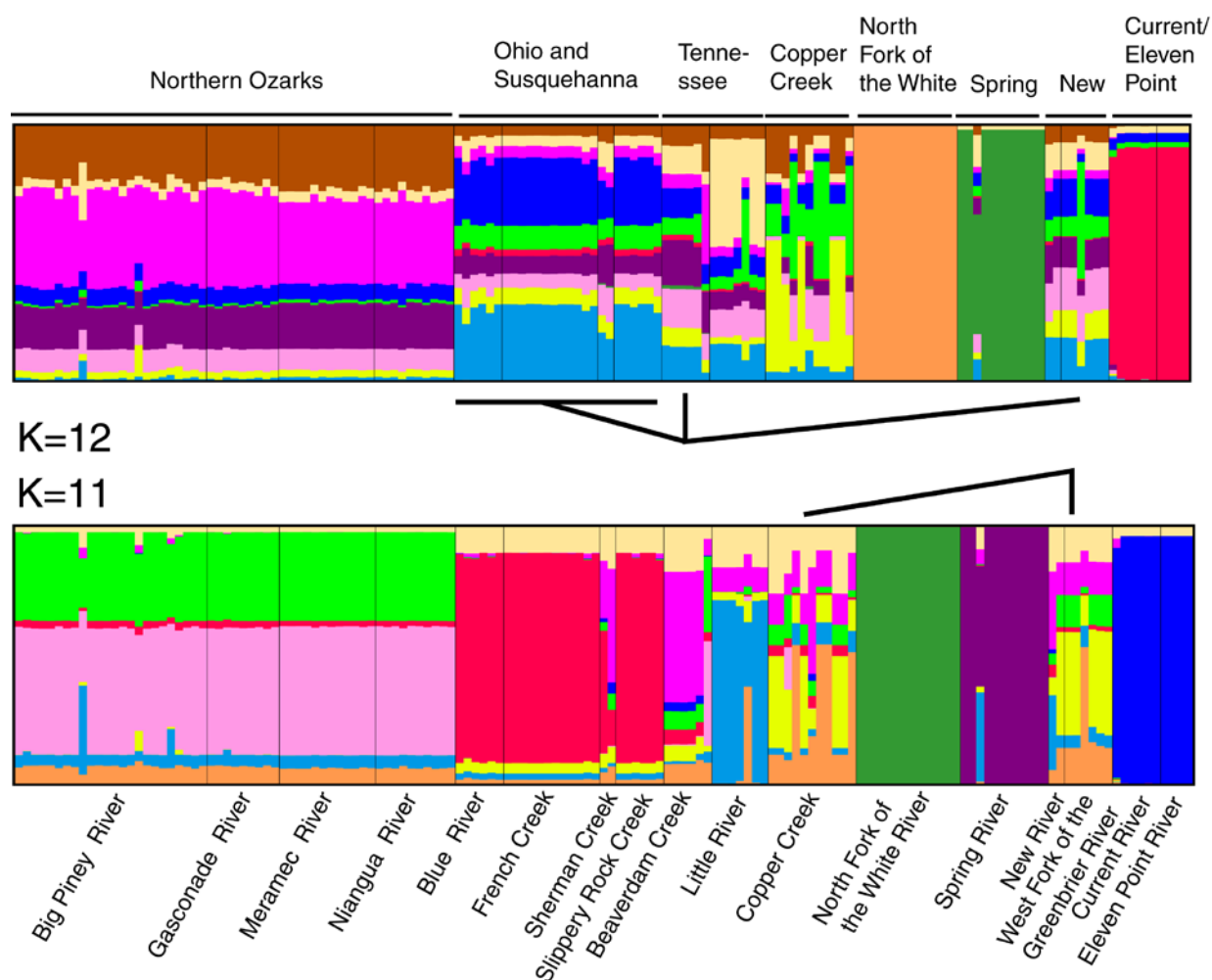


Figure 7. A comparison of the clusters estimated with STRUCTURE for $K=12$ vs. $K = 11$. Interpretation of the charts is as in Fig. 5. Colors cannot be compared across charts, because color assignment is independent in each analysis.

Table 1. Sampling locations (abbreviation), major drainages into which the rivers flow, and hellbender subspecies designation. All drainages eventually flow into the Mississippi River except Sherman Creek (Susquehanna) which terminates in the Chesapeake Bay (see Fig.1). Spacing between groups of rivers separates groups that drain into different major rivers prior to entering the Mississippi River.

Sampled River	Major Drainage	Subspecies
Gasconade River (GR)	Missouri	<i>alleganiensis</i>
Big Piney River (BP)	Gasconade-Missouri	<i>alleganiensis</i>
Niangua River (NR)	Osage-Missouri	<i>alleganiensis</i>
Meramec River (MR)	Mississippi	
Little River (LR)	Tennessee	<i>alleganiensis</i>
Beaverdam Creek (BC)	Holston-Tennessee	<i>alleganiensis</i>
Copper Creek (CC)	Clinch-Tennessee	<i>alleganiensis</i>
Blue River (BR)	Ohio	<i>alleganiensis</i>
French Creek (FC)	Allegheny-Ohio	<i>alleganiensis</i>
Slippery Rock Creek (SRC)	Connoquenessing-Ohio	<i>alleganiensis</i>
New River (NR)	Kanawha-Ohio	<i>alleganiensis</i>
West Fork of the Greenbrier River(WFGR)	New-Kanawha-Ohio	<i>alleganiensis</i>
Sherman Creek (SC)	Susquehanna	<i>alleganiensis</i>
North Fork of the White River (NFW)	White	<i>bishopi</i>
Spring River (SR)	Black-White	<i>bishopi</i>
Current River(CR)	Black-White	<i>bishopi</i>
Eleven Point River (EPR)	Black-White	<i>bishopi</i>

Table 2. Primer sequences for mitochondrial genes sequenced for hellbenders

Primer Name	Primer sequence (5'-3')	Reference
cytb		
tRNA-Glucine	TTGTATTCAACTATAAAAAC	Steele et al. (2005)
SAL-R1	ACTTAACCTCCTGTTGGTCA	
Cytb-CA-F1	TTCATTTATTGACCTACCAACC	
Cytb-CA-R1	GATAATTGACACTAAGGCTCAG	
COI		
LCO 1490	GGTCAACAAATCATAAAGATATTGG	Folmer et al. (1994)
HCO 2198	TAAACTTCAGGGTGACCAAAAAATCA	Folmer et al. (1994)
COI-CA-F1	TTAAGCTTATTAATCCGAGCAG	
COI-CA-R1	TGGCTGATGTGAAATAAGCTCG	
ND4		
ND4-Nex-H	TGGGGGCTACGGCATAATAC	Hardy et al. (2002)
ND4-Nex-L	CCAATGGATGAACTATTATCCT	Hardy et al. (2002)
ND4-Sal-F1	AAAATACCACTCTATGGTGC	
ND4-Sal-R1	G TTCATAACTTTCAC TTGGA	

Table 3. Primer sequences (5' – 3') with references and GenBank Accession numbers for the original locus

Locus	Primer	Sequence (5' – 3')	Reference	GenBank Accession no.
β-crystallin	CRYBA1Ls	CGCCTGATGTCTTTCCGCC	Dolman and Philips(2004)	AY508951 –
	CRYBA2Ls	CCAATGAAGTTCTCTTTCTCAA		AY508952
Myosin heavy chain	MYH2-F10	TGCACTCACAGAACACCAG	This study	AY508930
	MYH2-R	TGGTGTCTGCTCCTTCTT		
Steel	STI-CA-F1	ACGCCAGGTATGCCGAGG	This study	
	STI-CA-R2	GTGATGTTAGGAGTCAGTGC		
5S	5S-F1	GGGAGACTGCCTGGAATAC	This study	AB066111 –
	5S-R1	CGCATTCAAGGTGGTATGG		
eif2s1	Euktrans_x4	GGTCTGAGNTGTTCCACAGA	This study	EH168568
	Euktrans_x5	CNTANCGAGGAGGAGCAATC		
RPLP0	Ribpro_x1	CAGGGAAGACAGGGCTACNT	This study	EG018546
	Ribpro_x2	CCCCACAATGAAGCATT		
Ef-1α	Elofact_x6	CCNATGTGNGTGGAGAGCTT	This study	ES272822
	Elofact_x7	CATGTCACGGACAGCAAAAC		
GATA	GATAF1	AGGGATCCCCTGACAGAAGT	This study	
	GATAR	GCTTTGACTGGGCCATTCTA		

Table 4. Characterization of sequenced nuclear loci for *Cryptobranchus alleganiensis*

Locus	Product size	Polymorphic nucleotide sites	Indels	Population successfully amplified
β-crystallin	186	0	0	BP, CC, NFW*, SR*
Myosin heavy chain	254	1	0	BP, CR*, FC
Steel	388	10	2 (2, 5)	BP, CC, CR*, EPR*, GR, LR, NewR, NFW*, SC, SR*
5S	278	9	0	BC, BP, CC, CR*, EPR*, FC, LR, NewR, NFW, SR*, SRC
eif2s1	422	0	0	BR, EPR*, NFW*, SRC
RPLP0	248	0	0	FC, NFW*, SR*, SRC
Ef-1α	296	0	0	NFW*, SR*, SRC
GATA	659	6	1 (2)	BP, CC, CR*, NewR, SRC, SR*

Approximate size of allele in base pairs; the number of polymorphic sites; the number of sites with insertions/deletions (indels) and size of indels in base pairs in parentheses; population in which the locus was amplified: Beaverdam Creek (BC), Big Piney (BP), Copper Creek (CC), Current River (CR), Eleven Point River (EPR), French Creek (FC), Gasconade River (GR), Little River (LR), New River (NewR), North Fork of the White (NFW), Sherman Creek (SC), Spring River (SR), Slippery Rock Creek (SRC). Localities for *C. a. bishopi* are denoted by an asterisk (*).

Table 5. Haplotype distribution among sampling localities for three separate mitochondrial genes and 3-locus composite haplotypes. N = number of individuals sequenced for corresponding gene. Sample size for the three locus haplotypes is the minimum sample size for the three individual loci comprising the haplotype.

<i>River</i>	<i>N</i>	<i>Cytb haplo- type</i>	<i>N</i>	<i>COI haplo- type</i>	<i>N</i>	<i>ND4 haplo- type</i>	<i>3-locus haplotype Name</i>	<i>3-locus composite haplotype</i>
Big Piney River	6	4	13	2	3	3	BP_GR-1	4_2_3
	2	6	1	3	2	4	BP_GR-2	5_2_3
	4	5					BP-1	6_3_4
Gasconade River	3	5	4	2	3	3	BP_GR-2	5_2_3
	4	4	2	9	2	12	BP_GR-1	4_2_3
	2	6					GR-1	6_9_12
Niangua River	4	6	5	17	3	3	NR-1	21_17_3
	1	21					NR-2	6_17_3
Meramec River	8	16	7	12	6	3	MR-1	17_12_3
	2	17	1	13			MR-2	16_13_3

Beaverdam Creek	4	1	6	1	2	1	BC-1	2_1_2
	1	3			1	2	BC-2	1_1_1
	1	2					BC-3	3_1_1
Little River	5	13	5	1	2	1	LR-1	14_10_1
	1	15	1	11	1	13	LR-2	13_1_1
	1	14	1	10			LR-3	13_1_13
Copper Creek	9	8	8	5	3	7	CC-1	8_5_7
Slippery Rock Creek	1	25	3	4	1	10	SRC-1	24_4_6
	3	7	1	7	4	6	SRC_BR-1	7_4_6
	1	24	1	19			SRC-2	7_19_6
							SRC-3	25_7_10
New River	2	20	1	14	1	14	NewR-1	20_14_14
			1	15	1	15	NewR-2	20_15_15
Blue River	4	7	4	4	1	5	SRC_BR-1	7_4_6

					1	6	BR-1	7_4_5
French Creek	4	11	5	4	4	10	FC_SC-1	11_7_10
	1	12	3	7	3	6	FC-2	12_4_6
			1	8	2	11	FC_4	11_8_10
							FC-3	11_4_10
Sherman Creek	2	11	1	7	1	10	FC_SC-1	11_7_10
North Fork of the White River	4	19	6	16	2	16	NFW-1	19_16_16
	2	18					NFW-2	18_16_16
Spring River	1	22	1	18	6	17	SR-1	22_18_17
	1	23					SR-2	23_18_17
Current River	2	9	6	6	2	9	CR-1	9_6_9
					1	8		
Eleven Point River	1	10	3	3	3	9	EPR-1	10_3_9
	1	26						

Table 6. Minimum % sequence divergence (maximum likelihood distance) among populations for 3-locus haplotypes.

	SC	BR	FC	SRC	New	BP	GR	MR	NR	BC	LR	CC	NFW	SR	CR	EPR
Sherman Creek	-															
Blue River	0.2	-														
French Creek	0.0	0.1	-													
Slippery Rock Ck.	0.1	0.0	0.1	-												
New River	5.7	5.5	5.4	5.4	-											
Big Piney River	0.7	1.0	0.7	0.8	5.2	-										
Gasconade River	0.7	1.0	0.7	0.8	5.2	0.0	-									
Meramec River	0.8	1.1	0.8	0.9	5.2	0.2	0.2	-								
Niangua River	0.7	1.1	0.7	0.8	5.3	0.1	0.1	0.1	-							
Beaverdam Creek	3.2	3.2	3.0	3.1	4.7	3.3	3.3	3.3	3.4	-						
Little River	3.2	3.3	3.1	3.2	5.2	3.4	3.4	3.4	3.5	0.1	-					
Copper Creek	3.5	3.5	3.4	3.4	4.5	3.7	3.7	3.6	3.7	3.1	3.2	-				
North Fork White	4.6	4.4	4.3	4.4	5.0	4.2	4.2	4.2	4.3	3.6	3.7	3.9	-			
Spring River	4.3	3.9	3.8	3.9	4.5	3.9	3.9	3.9	3.9	3.3	3.5	3.2	1.7	-		
Current River	5.7	5.8	5.5	5.6	4.3	5.5	5.5	5.5	5.6	5.0	5.1	4.9	5.3	4.8	-	
Eleven Point River	5.3	5.6	5.3	5.4	3.8	5.3	5.3	5.2	5.4	4.4	4.6	4.7	4.8	4.4	0.3	-

Table 7. Number of alleles per locus for microsatellite loci.

Locus	River																
	BP	GR	MR	NR	BR	FC	SC	SRC	BC	LR	CC	NFW	SR	NewR	WFGR	CR	EPR
GATA	7	5	5	6	5	7	3	5	5	6	5	6	9	4	8	10	3
CRAL4	2	2	1	3	5	7	1	4	6	6	3	2	1	4	7	2	2
CRAL9	4	2	1	2	4	3	2	4	3	4	6	1	3	3	4	2	1
CRAL13	3	3	3	2	4	3	2	5	4	5	5	2	1	2	3	2	1
Mean	4.0	3.0	2.0	3.2	4.5	5.0	1.7	4.5	4.5	5.2	4.7	2.5	3.0	3.2	5.5	4.0	1.2

Table 8. Expected heterozygosities for microsatellite loci.

Locus	River																
	BP	GR	MR	NR	BR	FC	SC	SRC	BC	LR	CC	NFW	SR	NewR	WFGR	CR	EPR
GATA	0.733	0.804	0.591	0.800	0.864	0.859	0.833	0.848	0.788	0.835	0.740	0.748	0.900	1.000	0.924	0.970	0.679
CRAL4	0.401	0.209	0.000	0.574	0.788	0.736	0.000	0.636	0.848	0.864	0.515	0.369	0.000	1.000	0.879	0.167	0.250
CRAL9	0.370	0.294	0.000	0.100	0.712	0.236	0.500	0.636	0.591	0.659	0.801	0.000	0.255	0.833	0.682	0.167	0.000
CRAL13	0.606	0.699	0.554	0.505	0.561	0.627	0.667	0.667	0.712	0.727	0.784	0.148	0.000	0.667	0.682	0.167	0.000
Mean	0.528	0.502	0.286	0.495	0.731	0.614	0.500	0.697	0.735	0.771	0.710	0.316	0.289	0.875	0.792	0.367	0.232

Appendix 1. Table of specimen genotypes for 4 microsatellite alleles. Each individual is listed twice enable the listing of both alleles at each locus. Individuals are listed in the same order as in the STRUCTURE output graphs.

SPECIMEN NUMBER	POPULATION	GATA	CRAL4	CRAL9	CRAL13
553	BP	148	169	137	187
553	BP	148	169	137	191
562	BP	148	169	133	187
562	BP	152	169	133	187
564	BP	144	169	133	187
564	BP	152	169	137	187
593	BP	152	173	137	187
593	BP	152	173	137	197
594	BP	144	---	127	187
594	BP	152	---	137	191
595	BP	148	169	137	187
595	BP	156	169	137	191
596	BP	148	169	133	187
596	BP	152	169	137	191
597	BP	144	---	137	187
597	BP	168	---	137	191
598	BP	148	---	---	187
598	BP	148	---	---	191
599	BP	144	169	137	187
599	BP	152	173	137	191
603	BP	144	169	137	187
603	BP	144	169	137	197
604	BP	152	169	137	187
604	BP	152	173	137	191
607	BP	152	169	133	187
607	BP	156	169	137	197
608	BP	136	173	137	187
608	BP	148	173	137	191
610	BP	152	169	137	187
610	BP	156	173	137	197
670	BP	152	---	127	197
670	BP	152	---	137	197
671	BP	148	169	137	187
671	BP	152	173	137	197
673	BP	144	---	137	187
673	BP	148	---	137	191
677	BP	152	169	137	187
677	BP	164	169	137	191
710	BP	152	---	133	187
710	BP	156	---	133	187

711	BP	152	169	135	187
711	BP	152	169	137	197
719	BP	148	169	137	187
719	BP	152	169	137	191
769	BP	156	---	137	187
769	BP	156	---	137	191
945	BP	152	169	137	187
945	BP	152	173	137	191
753	GR	156	169	133	187
753	GR	160	169	137	197
754	GR	144	169	137	197
754	GR	144	173	137	197
755	GR	148	169	133	187
755	GR	148	169	133	191
756	GR	148	169	137	187
756	GR	152	169	137	191
757	GR	152	169	137	187
757	GR	152	173	137	191
760	GR	144	169	137	187
760	GR	148	169	137	191
761	GR	152	169	137	187
761	GR	156	169	137	191
1131	GR	152	169	137	197
1131	GR	156	169	137	197
1136	GR	152	169	137	187
1136	GR	156	169	137	191
1106	MR	156	169	137	187
1106	MR	160	169	137	191
1118	MR	156	---	137	187
1118	MR	156	---	137	191
1119	MR	156	169	137	187
1119	MR	160	169	137	191
1120	MR	156	169	137	187
1120	MR	160	169	137	191
1144	MR	144	169	137	187
1144	MR	156	169	137	191
1145	MR	156	169	137	187
1145	MR	156	169	137	191
1146	MR	152	169	137	187
1146	MR	160	169	137	187
1147	MR	148	169	137	187
1147	MR	156	169	137	187
1427	MR	156	169	137	187
1427	MR	156	169	137	191
1429	MR	156	169	137	191

1429	MR	156	169	137	191
1430	MR	152	169	137	187
1430	MR	156	169	137	191
1431	MR	148	169	137	187
1431	MR	156	169	137	193
566	NR	156	169	137	187
566	NR	164	173	137	191
568	NR	156	169	131	187
568	NR	156	173	137	191
569	NR	156	169	137	187
569	NR	160	173	137	187
570	NR	152	169	137	187
570	NR	152	173	137	191
571	NR	148	169	137	187
571	NR	164	175	137	187
665	NR	156	169	137	187
665	NR	160	169	137	191
666	NR	148	173	137	187
666	NR	148	173	137	191
678	NR	160	169	137	187
678	NR	164	173	137	191
679	NR	156	173	137	187
679	NR	156	173	137	191
681	NR	156	169	137	187
681	NR	168	173	137	191
1485	BR	148	173	135	197
1485	BR	156	175	135	197
1486	BR	140	175	131	195
1486	BR	144	185	139	199
1487	BR	152	165	139	197
1487	BR	152	175	139	197
1488	BR	144	171	139	195
1488	BR	148	173	141	197
1489	BR	144	175	139	197
1489	BR	156	185	139	203
1490	BR	140	173	141	197
1490	BR	152	175	141	197
1491	FC	148	173	139	193
1491	FC	156	175	139	197
1492	FC	140	175	139	197
1492	FC	160	189	139	201
1493	FC	152	175	139	197
1493	FC	152	175	141	197
1494	FC	148	173	139	197
1494	FC	148	173	139	201

1495	FC	140	171	139	193
1495	FC	152	175	139	201
1496	FC	144	175	139	197
1496	FC	152	185	141	201
1497	FC	144	173	139	193
1497	FC	156	189	139	197
1498	FC	144	169	139	193
1498	FC	148	175	139	197
1499	FC	148	175	139	197
1499	FC	164	177	139	201
1500	FC	144	173	139	197
1500	FC	156	173	139	201
1501	FC	144	173	129	193
1501	FC	152	173	139	197
1502	FC	140	173	139	197
1502	FC	156	175	139	197
1384	SC	152	173	139	199
1384	SC	156	173	139	199
1385	SC	168	173	139	197
1385	SC	168	173	143	197
951	SRC	148	173	139	197
951	SRC	152	175	141	197
1505	SRC	156	173	139	195
1505	SRC	164	189	141	197
1506	SRC	152	173	139	197
1506	SRC	156	175	139	197
1507	SRC	148	171	139	197
1507	SRC	152	173	139	201
1508	SRC	144	173	135	193
1508	SRC	148	173	141	201
1484	SRC	144	173	129	197
1484	SRC	156	175	139	203
613	BC	152	173	129	189
613	BC	156	173	129	199
614	BC	156	173	133	197
614	BC	164	175	133	199
615	BC	156	169	133	189
615	BC	164	175	133	197
616	BC	136	169	129	197
616	BC	156	173	133	197
619	BC	152	177	129	189
619	BC	168	193	133	197
620	BC	152	169	131	191
620	BC	156	201	133	197
621	LR	160	161	123	187

621	LR	160	197	133	191
622	LR	148	173	123	187
622	LR	148	199	147	199
623	LR	148	161	123	187
623	LR	160	175	133	197
624	LR	152	177	123	187
624	LR	152	197	123	193
630	LR	144	---	123	---
630	LR	152	---	133	---
631	LR	152	161	129	187
631	LR	164	199	133	191
632	LR	148	161	123	187
632	LR	168	173	133	191
1218	CC	152	167	131	197
1218	CC	152	167	131	203
1369	CC	152	167	131	197
1369	CC	164	167	135	205
1370	CC	164	167	137	---
1370	CC	164	169	137	---
1371	CC	148	167	129	195
1371	CC	152	169	133	203
1372	CC	152	167	131	197
1372	CC	164	181	131	197
1373	CC	156	167	129	---
1373	CC	164	169	133	---
1374	CC	144	167	129	195
1374	CC	152	169	133	205
1375	CC	144	167	129	195
1375	CC	152	169	133	205
1376	CC	148	167	131	197
1376	CC	164	169	131	205
1377	CC	164	167	131	197
1377	CC	164	169	135	203
1378	CC	144	167	127	197
1378	CC	152	167	133	201
638	NFW	144	163	127	199
638	NFW	152	163	127	199
639	NFW	148	163	127	199
639	NFW	148	163	127	199
640	NFW	140	163	127	199
640	NFW	144	165	127	199
642	NFW	148	163	127	199
642	NFW	160	163	127	199
643	NFW	144	163	127	199
643	NFW	164	163	127	199

645	NFW	148	163	127	199
645	NFW	160	163	127	199
649	NFW	144	163	127	199
649	NFW	148	165	127	199
1513	NFW	140	163	127	195
1513	NFW	160	165	127	199
1514	NFW	152	163	127	199
1514	NFW	160	163	127	199
1530	NFW	148	163	127	199
1530	NFW	148	165	127	199
1537	NFW	148	163	127	199
1537	NFW	152	165	127	199
1542	NFW	148	163	127	195
1542	NFW	148	163	127	199
100b	NFW	148	163	127	199
100b	NFW	148	165	127	199
592	SR	156	161	119	195
592	SR	180	161	119	195
1148	SR	172	161	119	195
1148	SR	204	161	119	195
1154	SR	---	161	125	---
1154	SR	---	161	125	---
1155	SR	160	161	119	195
1155	SR	172	161	119	195
1156	SR	172	161	119	195
1156	SR	204	161	119	195
1157	SR	168	161	119	195
1157	SR	168	161	119	195
1158	SR	172	161	119	195
1158	SR	192	161	119	195
1159	SR	160	161	119	195
1159	SR	176	161	119	195
1160	SR	168	161	119	195
1160	SR	172	161	119	195
1161	SR	196	161	119	195
1161	SR	204	161	127	195
1162	SR	180	161	119	195
1162	SR	192	161	119	195
1216	NEW	156	177	125	193
1216	NEW	168	185	125	199
1217	NEW	148	167	129	193
1217	NEW	160	179	131	199
WV1	WFG	172	179	129	193
WV1	WFG	181	179	131	197
WV2	WFG	164	177	131	193

WV2	WFGR	176	187	131	193
WV3	WFGR	144	183	129	195
WV3	WFGR	152	193	129	195
WV4	WFGR	160	167	129	193
WV4	WFGR	164	197	129	195
WV5	WFGR	148	187	129	193
WV5	WFGR	152	193	137	197
WV6	WFGR	152	187	131	193
WV6	WFGR	160	187	135	197
1135	CR	168	173	127	197
1135	CR	184	175	127	197
1169	CR	208	175	127	195
1169	CR	212	175	127	197
1170	CR	164	175	127	197
1170	CR	196	175	127	197
1171	CR	196	175	127	197
1171	CR	204	175	127	197
1172	CR	156	175	119	197
1172	CR	160	175	127	197
1173	CR	164	175	127	197
1173	CR	188	175	127	197
81	EPR	164	175	127	197
81	EPR	172	175	127	197
83	EPR	164	175	127	197
83	EPR	172	175	127	197
88	EPR	160	175	127	197
88	EPR	164	175	127	197
92	EPR	164	175	127	197
92	EPR	172	177	127	197

Appendix 2. Cluster membership probabilities for each population for STRUCTURE results (K=8). N=number in sample. Numbers in the header row refer to cluster number, which is shown on the STRUCTURE graphical output by placing each number on its corresponding color. Values in the cells are averages of the individual probabilities. The value of 0 was used instead of 0.0000 to improve readability.

Pop	CLUSTER								N
	1	2	3	4	5	6	7	8	
BP	0	0.0192	0.0369	0	0	0.0003	0.6388	0.3048	24
GR	0	0.0126	0.0159	0	0	0	0.6665	0.3050	9
MR	0	0.0168	0	0	0	0	0.5893	0.3939	12
NR	0	0.0162	0.0004	0	0	0	0.5980	0.3854	10
BR	0.9983	0.0002	0.0010	0	0	0	0.0001	0.0004	6
FC	0.9998	0	0	0	0	0	0	0.0002	12
SC	0.8706	0.0104	0.0529	0	0	0.0392	0.0040	0.0229	2
SRC	0.9991	0.0002	0	0	0	0	0.0001	0.0006	6
BC	0.0035	0.0369	0.7082	0	0	0.0003	0.1099	0.1412	6
LR	0	0.0151	0.9494	0	0	0	0.0128	0.0227	7
CC	0	0.7025	0.0218	0	0	0	0.0333	0.2424	11
NFW	0	0	0	1.0000	0	0	0	0	13
SR	0	0	0.0422	0	0.9578	0	0	0	11
NewR	0.1049	0.4716	0.3704	0	0.0017	0	0.0020	0.0494	2
WFGR	0	0.8874	0.0073	0	0	0	0.0031	0.1022	6
CR	0.0007	0.0004	0	0	0	0.9982	0.0001	0.0006	6
EPR	0	0	0	0	0	1.0000	0	0	4

